Detecting tones in complex auditory scenes

Giani, Anette S.; Belardinelli, Paolo; Ortiz, Erick; Kleiner, Mario; Noppeney, Uta

DOI:
10.1016/j.neuroimage.2015.07.001

Detecting tones in complex auditory scenes

Anette S. Giani, Paolo Belardinelli, Erick Ortiz, Mario Kleiner, Uta Noppeney

PII: S1053-8119(15)00608-4
DOI: doi: 10.1016/j.neuroimage.2015.07.001
Reference: YNIMG 12390

To appear in: *NeuroImage*

Received date: 18 February 2014
Accepted date: 1 July 2015

Please cite this article as: Giani, Anette S., Belardinelli, Paolo, Ortiz, Erick, Kleiner, Mario, Noppeney, Uta, Detecting tones in complex auditory scenes, *NeuroImage* (2015), doi: 10.1016/j.neuroimage.2015.07.001

Detecting tones in complex auditory scenes

Abbreviated title: From complex scenes to auditory awareness

Anette S. Giani[1†], Paolo Belardinelli[2†], Erick Ortiz[2], Mario Kleiner[1] & Uta Noppeney[1,3]

[1] Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

[2] Functional and Restorative Neurosurgery, University Hospital Tübingen, 72076 Tübingen, Germany

[3] Centre for Computational Neuroscience and Cognitive Robotics, Department of Psychology, University of Birmingham, B15 2TT Birmingham, UK

[†] These authors contributed equally to this work.

Corresponding author:
Paolo Belardinelli
Functional and Restorative Neurosurgery
University Hospital Tübingen
Eberhard Karls University Tübingen
Otfried-Mueller-Str.45
D-72076 Tuebingen
paolo.belardinelli@uni-tuebingen.de

Number of *pages*: 41, *figures*: 4, *tables*: 4
Number of words: *Abstract:* 249, *Introduction*: 560, *Discussion*: 1612

Acknowledgments:
This work was supported by the Max Planck Society and the European Research Council (ERC-multsens). We thank Jürgen Dax for technical support and the Cognitive Neuroimaging Group. The authors declare no competing financial interests.

Abstract

In everyday life, our auditory system is bombarded with many signals in complex auditory scenes. Limited processing capacities allow only a fraction of these signals to enter perceptual awareness. This magnetoencephalography (MEG) study used informational masking to identify the neural mechanisms that enable auditory awareness. On each trial, participants indicated whether they detected a pair of sequentially presented tones (i.e., the target) that were embedded within a multi-tone background.

We analysed MEG activity for 'hits' and 'misses', separately for the first and second tone within a target pair. Comparing physically identical stimuli that were detected or missed provided insights into the neural processes underlying auditory awareness. While the first tone within a target elicited a stronger early P50m on hit trials, only the second tone evoked a negativity at 150ms, which may index segregation of the tone pair from the multi-tone background. Notably, a later sustained deflection peaking around 300 and 500 ms (P300m) was the only component that was significantly amplified for both tones, when they were detected pointing towards its key role in perceptual awareness.

Additional Dynamic Causal Modelling analyses indicated that the negativity at 150 ms underlying auditory stream segregation is mediated predominantly via changes in intrinsic connectivity within auditory cortices. By contrast, the later P300m response as a signature of perceptual awareness relies on interactions between parietal and auditory cortices.

In conclusion, our results suggest that successful detection and hence auditory awareness of a two-tone pair within complex auditory scenes relies on recurrent processing between auditory and higher-order parietal cortices.

1. Introduction

In our natural environment, our senses are exposed to a constant influx of sensory signals, only a fraction of which enters our perceptual awareness (Dehaene et al., 2014; Koch and Tsuchiya, 2007). Current influential theories suggest that perceptual awareness relies on (i) recurrent neural processing via long-distance connections between low and higher order sensory areas (Gutschalk et al., 2008; Lamme and Roelfsema, 2000; Ro et al., 2003) or (ii) interactions between sensory and frontoparietal areas (Baars, 1993; Dehaene et al., 2003). In particular, interactions with the frontoparietal network are thought to be critical for sharing information widely within a global neuronal workspace (reviewed in: Dehaene et al., 2006). These theories have been developed primarily based on studies of visual awareness. More recently, studies have also started investigating whether similar principles apply to perceptual awareness in other sensory modalities such as audition (Allen et al., 2000; Bekinschtein et al., 2009; Diekhof et al., 2009; Sadaghiani et al., 2009).

In the auditory system, the brain needs to break a complex sound wave into streams that emanate from independent sound sources and can then independently tracked in perceptual awareness (Bregman, 1990). Auditory awareness of objects in complex auditory scenes thus inherently relies on figure ground segregation.

Informational masking has proven a powerful paradigm to study how auditory streams are segmented and enter perceptual awareness. In informational masking paradigms, participants need to detect a target sequence that is defined by the repetition of two (or more) identical tones within a multi-tone mask (Neff and Green, 1987). Recent MEG studies of informational masking have demonstrated that auditory steady-state signals generated in primary auditory cortices and early responses such as the P50m did not differentiate between detected and undetected target sequences (Konigs and Gutschalk, 2012; Wiegand and Gutschalk, 2012). By contrast, the so-called ‘awareness-related negativity’, a cortical response peaking at about 150 ms, emerged

only for detected target sequences (Gutschalk et al., 2008). These findings have led to the conjecture that auditory awareness is associated with neural processing in secondary auditory cortices. Yet, related fMRI/EEG studies focusing on ABA streaming, attentional blink or figure ground segregation associated parietal cortices with auditory stream segregation or temporal figure ground segregation (Cusack, 2005; Teki et al., 2011) and/or perceptual awareness (Bekinschtein et al., 2009; Del Cul et al., 2007; Haynes et al., 2005; Sergent et al., 2005). These discrepancies may be explained by the fact that the informational masking studies focused selectively on responses in auditory cortices or the first 400 ms poststimulus (Gutschalk et al., 2008; Wiegand and Gutschalk, 2012). Furthermore previous analyses have not characterized the build-up of auditory stream segregation by directly comparing target tones that were detected vs. undetected individually for each position.

Using an informational masking paradigm, the current MEG study presented participants with a pair of tones (i.e. target) embedded in a multi-tone mask. First, we analysed the evoked responses separately for first and second tones within a target pair in time windows extending to 600 ms poststimulus. This enabled us to define the complete neural processing cascade that is associated with successful detection of a target pair within complex auditory scenes. Second, we investigated whether successful target detection relies on recurrent neural processing between sensory and frontoparietal areas (Baars, 1993; Dehaene et al., 2003). We combined Dynamic Causal Modelling (Kiebel et al., 2009) and Bayesian Model Comparison to evaluate the contributions of intrinsic, forward and backward connections to auditory stream segregation and successful detection of target pairs in early [<300ms] and early+late [<600ms] time windows.

2. Materials and Methods

2.1 Participants

After giving informed consent, 21 healthy young adults participated in this study (12 females, 20 right-handed, mean age (± standard deviation): 26.2 ± 4.04 years, range: 20-36 years). All of them reported normal hearing and had normal or corrected-to-normal vision. Data from one female participant was discarded from the analysis due to excessive eye blinks. The study was approved by the local ethics review board of the University of Tübingen.

2.2 Experimental design

On each trial, participants were instructed to detect a pair of tones that were embedded within a multi-tone mask on 2/3 of the trials (target present trials) and absent in the remaining 1/3 trials (target absent or catch trials) (**Figure 1A**). The target pair of tones was presented with a fixed stimulus onset asynchrony in a protected frequency zone (see stimuli section). In additional trials, visual gratings were presented alone or together with the auditory targets; however, these trials are not included or discussed in this report.

Thus, this report focuses only on the auditory target present trials that were categorized post hoc into ‘hits’ and ‘misses’ depending on whether participants successfully detected the target pair of two tones. Comparing hits and misses to identical auditory stimuli enabled us to assess the neural processes that are associated with successful target detection. Critically, the target was defined as a pair of two sequentially presented tones. Hence, the categorization as hit or miss pertains always to both tones. Because the two tones were presented sequentially, the MEG responses for a hit or miss trial could be analysed separately for the first and second tone. Thus, our paradigm enabled us to compare hits vs. misses separately for the first and second

tone within a target pair and characterize the cascade of processes associated with segmentation of a pair of tones from a complex auditory scene. Moreover, as successful task performance requires participants to detect both tones within the target, we reasoned that neural signatures of auditory awareness should be present for both target tones when they are successfully detected. However, we acknowledge that the first tone may also remain initially subliminal and only be boosted into participants' awareness by the presentation of the 2nd tone i.e. 1050 ms after the onset of the 1st tone (e.g. Backer and Alain, 2012; Sergent et al., 2013). Moreover, while it is a standard procedure to define perceptual awareness based on participants' subjective report (e.g. detection responses as in our study), we acknowledge that this approach is not free of flaws and interpretational ambiguities. Most importantly, it cannot dissociate perceptual from more decisional processes responsible for participants making a hit response. Misses may differ from hits in sensory noise leading to less reliable or weaker sensory representations. Conversely, participants may set a higher decisional criterion on miss trials, so that identical sensory representations may be judged as noise rather than signal. Future experiments explicitly manipulating participants' decisional criterion for instance using cuing paradigms where the cue predicts the probability of target presence may help us further to disentangle perceptual from more decisional processes in auditory detection during informational masking.

In summary, we compared the MEG activity for hits vs. misses separately for the 1st and 2nd tone within a target pair (**Figure 1B**). Please note that even though we will analyse and report the MEG responses to the 1st and 2nd tones within a target pair separately, participants needed to detect a target pair that consists of two tones. Hence, a particular trial (i.e. with a target pair of two tones) was classified as a hit or a miss. Thus, this experiment could be analysed in a 2 (detected vs. not detected) x 2 (tone 1vs. tone 2) factorial fashion.

Figure 1 about here

2.3 Stimuli

All target and masking tones were amplitude modulated tones with a duration of 300 ms and a carrier frequency selected from a set of 26 frequencies. These frequencies were equally spaced on a logarithmic scale ranging from 200 to 5065 Hz.

On each trial, one single carrier frequency was selected commonly for both of the tones within a target pair from a set of five possible target frequencies: 1222, 1583, 2049, 2654 and 3437 Hz. This target frequency and additional 3 frequencies above and below the target frequency were then protected from being used as masking tones (i.e. protected frequency zone). Hence, the multi-tone, informational mask included only tones of the remaining 19 frequencies. To maximize variability of the masking tones, these frequencies varied around one estimated rectangular bandwidth [ERB = 24.7 x (4.37 x fc + 1)] with fc = centre frequency in kHz (Gutschalk et al., 2008; Moore, 1995).

The SOA of the two tones within a target pair was held constant at 1050 ms throughout the entire experiment. The SOAs of the masking tones were randomized within each frequency band between 550 and 1550ms (mean 1050 ms), excluding a protected region of 850 to 1250 ms centred on the fixed SOA of 1050 ms of two tones within a target pair (i.e. protected SOA zone). The SOAs were sampled independently for each trial and subject. On average 19 masker tones were presented in each 1050 ms window (i.e. 1 masker tone/ 55 ms).

Similar to previous informational masking studies (Gutschalk et al., 2008; Kidd et al., 1994; Kidd et al., 2003; Konigs and Gutschalk, 2012; Micheyl et al., 2007) participants could use predominantly two cues to segregate and detect the target pair of tones from the multi-tone background mask: (i) the target frequency was presented in a protected frequency zone and (ii) the fixed temporal interval between the two tones within a target pair was set in a protected

SOA zone (i.e. no masking tones were repeated with an SOA between 850 ms and 1250 ms). While the more subtle frequency cue was available already at the presentation of the first tone, the more prominent temporal SOA cue emerged only after the presentation of the 2nd tone and enabled segregation of the target pair from the multi-tone mask.

To evoke SSRs the target tones' amplitude was modulated sinusoidally at a rate of 40 Hz and a modulation depth of 100% which has consistently been shown to evoke robust auditory SSRs (for a review see: Picton et al., 2003). The masking tones' amplitude was modulated sinusoidally at rates of 32, 36, 44 or 48 Hz, thus enhancing the similarity between masking and target tones. To avoid clicking sounds, the masking tones were multiplied with 10 ms sinusoidal ramps at on- and offsets. This procedure was not needed for the target tones, because the 40 Hz AM modulation enabled a natural ramping for 300 ms tones. This is because 12 cycles of 25 ms naturally combined into 300 ms duration for 40 Hz AM tones. Target tones were presented at a mean loudness of 50.5 dB sensation level (SL), while masking tones level was, on average, 4 dB louder. Thus, the difference in sound amplitude between target and masking tones provided a third cue for identifying the softer target tones amongst masking tones.

All auditory stimuli were generated and controlled using MatLab, Psychtoolbox version 3.09 (revision 1754) (Brainard, 1997; Kleiner et al., 2007) running on an Apple MacBook Pro under Macintosh OS-X 10.6.7. Tones were digitized at a sampling rate of 44.8 kHz via the computer's internal HDA sound chip and presented binaurally via insert earphones (E-A-RTONE® 3A, Aero Company, USA). Precise on- and offsets of the stimuli were verified using photodiode and microphone recordings.

2.4 Schematic of the trial procedure

At the beginning of each trial there was a random delay period of 800 – 1300 ms (mean 1050 ms). Subsequently, on target present trials, the first tone within a target pair was played for a period of 300 ms. After a fixed stimulus onset asynchrony (SOA) of 1050 ms, the second tone of

the pair was presented. 1050 ms after the onset of the second tone the question: “Was there a target?” appeared on the screen.

A multi-tone mask and a central fixation cross were presented throughout the entire trial. Participants fixated the cross in the centre of the screen. In an auditory selective attention task, participants were asked to detect a pair of tones that were presented sequentially with the same carrier frequency and a fixed temporal interval of approximately one second. It was emphasized that they should only respond ‘yes’, if they detected both tones with confidence. This instruction ensured that participants applied a high decision criterion, which may potentially lead to ‘partially detected tones’ classified as misses. Participants were familiarized with the paradigm, the target and the masking tones in a prior training session, such that they could reliably detect the target pair of tones and were confident about their decisions. Participants indicated their response via a two choice key press with their right index or middle finger (order randomized) within a maximal response time of 2 seconds after the question appeared on the screen.

2.5 Experimental procedures

At the beginning of the experiment, participants’ detection thresholds were measured for pure tones with a carrier frequency corresponding to the 5 target frequencies and a set of standard frequencies: 250 500 1000 2000 and 4000 Hz based on the methods of limits. Briefly, participants were presented with series of tones that ascended or descended in 2 dB steps. They indicated for each tone whether or not they perceived a tone. The ascending and descending series were repeated twice and the average of the detection thresholds was selected for the main experiment. This process was repeated for the 5 target frequencies and the five additional standard frequencies.

These subject-specific thresholds were used to scale the sounds’ intensities separately for the masking and the target sounds to an equal level of loudness across different frequencies. Next,

participants were familiarized with the stimuli and task in a total of 2-5 short test sessions of the informational masking paradigm. Only during those training sessions, did participants receive visual feedback after each trial.

The experiment comprised 8 sessions, which were separated on two days. Each session included 240 trials, resulting in a total of 1920 trials. Hence, on each day the scanning time amounted to about 2 hours (including breaks). After each session the participant's head position was adjusted to fit the position of the first session as accurately as possible.

2.6 Data acquisition

Neuromagnetic data were recorded at 1171.88 Hz sampling frequency with a 275-channel whole-head MEG System (VSM, MedTech, Port Coquitlam, Canada; 275 axial gradiometers with 5 cm baseline and 29 reference channels) at the MEG Center Tübingen, Germany. Participants' head position was continuously monitored by three sensor coils attached to the nasion, and left and right pre-auricular (15 mm anterior to the left and right tragus) points of each subject. The positions of these coils, i.e. the fiducial points, were marked on the subject's skin. To measure eye movements and blinks, horizontal and vertical electrooculogram (EOG) were recorded from two pairs of bipolar electrodes.

A 3T Siemens Magnetom Tim Trio System (Siemens, Erlangen, Germany) at the MPI for Biological Cybernetics, Tübingen, Germany, was used to acquire high-resolution structural images (176 sagittal slices, TR = 2300 ms, TE = 2.98 ms, TI = 1100 ms, flip angle = 9°, FOV = 256 mm x 240 mm x 176 mm, voxel size = 1 mm x 1 mm x 1 mm). MR-markers that can be identified on the anatomical image were attached to the same fiducial points as described above to enable accurate co-registration of the anatomical MRI and the MEG data.

2.7 Data analysis

The MEG, MRI and behavioural data were pre-processed and analysed using statistical parametric mapping SPM8 (http://www.fil.ion.ucl.ac.uk/spm/; Wellcome Trust Centre of Neuroimaging, London, UK), fieldtrip (http://www.ru.nl/donders/fieldtrip) (Oostenveld et al., 2011) and MatLab 7 (MathWorks, Inc., Massachusetts, USA).

2.7.1 Sensor space: Pre-processing and statistics

For the sensor space analysis, we focused on event-related magnetic fields. Therefore, the continuous data were high-pass (cut-off: 1 Hz) and low-pass (cut-off: 30 Hz) filtered in forwards and reverse directions, using a 5^{th} order Butterworth digital filter. We were required to apply a filter of 1 Hz to our data in order to remove environmental low frequency noise. Hence, very low-frequency drifts and the 40 Hz SSR were removed. The filtered data were then down-sampled to 213.3 Hz and epoched into segments from -170 ms to 880 ms after the onset of the tone. The epoched data were baseline corrected by subtracting the activity averaged between -150 to -50 ms from all MEG channels. Noisy epochs (i.e. 3.9 % of all epochs) were rejected when the MEG signal exceeded 1.8 pT. Hence, eye movement, muscle and other short-lived artefacts were manually removed. Independent component analysis (ICA) was applied to correct for eye blink and heart beat artefacts. Eye blink and heartbeat-related components were identified based on visual inspection of component topographies and time-courses. In all datasets one single ICA component was related to eye blinks, while 1-2 (mean: 1.8) ICA components were related to heart beats. Finally, artefact-cleaned epochs were merged across sessions and averaged across trials to create event-related fields.

The linearly interpolated topography x time data were converted to 3D images (voxel size: 2.1 mm x 2.7 mm x 4.7 ms, image dimension: 64 x 64 x 214). The resulting images were smoothed in space and time, using an isotropic Gaussian Kernel of 12 mm/ms full-width at half maximum. At the random effects or between-subject level, for each subject 3D images were entered into several paired t-tests. Specifically, we performed the following three statistical comparisons:

(1) The effect of *target detection* on processing of the first auditory tone: Hit1 vs. Miss1.

(2) The effect of *target detection* on processing the second auditory tone: Hit2 vs. Miss2.

(3) The interaction between *target detection* and *tone-position within a target pair*: Hit1-Miss1 vs. Hit2-Miss2.

Based on our a priori hypotheses we restricted the analysis to the three time windows of interest (1) an early M_1 response (40-90 ms) (Boutros and Belger, 1999; Wiegand and Gutschalk, 2012), (2) the ARN (100-200ms) (Gutschalk et al., 2008; Konigs and Gutschalk, 2012; Wiegand and Gutschalk, 2012) and (3) the later long-latency M_3 (250-550 ms) (Ishii et al., 2009). Please note that we use the labels M_1 and M_3 in a purely descriptive fashion in order not to associate neural effects a priori with classical cognitive components. Instead, these labels simply to refer to the first and third components poststimulus. In the results section, we will then discuss and relate them to well-established components such as P50m or P300m based on scalp topography and timecourse.

The time windows' latencies were selected *a priori,* guided by previous studies and visual inspection of the mean time courses, pooled over all conditions and participants. This does not invalidate our inference as the mean activity across all conditions is orthogonal to our contrasts of interest. Unless otherwise stated, we report effects at $p < 0.05$ at the peak level corrected for multiple comparisons, within the entire interpolated scalp space and the time windows of interest, using random field theory.

2.7.2 Source space analysis: MRI processing, MEG-MRI coregistration and forward modelling

Structural MRI images were segmented and normalized to MNI space using unified segmentation (Ashburner and Friston, 2005). The inverse of this normalization transformation was employed to warp a template cortical mesh, i.e. a continuous tessellation of the cortex

(excluding cerebellum) with 20484 vertices, from MNI space to each subject's native space. The MEG data were projected onto each subject's MRI space by applying a rigid body coregistration using the fiducials as landmarks. As head model, we employed a single shell aligned with the inner skull. Lead fields were then computed for each vertex in the cortical mesh with each dipole oriented normally to that mesh.

2.7.3 Model inversion and source space statistics

Source localization was performed within a Bayesian framework using the Greedy Search (GS) algorithm implemented in SPM8 (version r4667), individually for each participant within a time-window from 0 to 600 ms. This time-window includes all windows of interest from our sensor space analysis. For each participant, the bandpass filtered (1-30 Hz) MEG data for all conditions were convolved with a Hanning window (used to down-weight baseline noise) and inverted together using 1024 patches per hemisphere (plus 1024 bilateral patches) of the cortical mesh.

Before inversion, the data were projected to a subspace of 73-102 (across-subjects mean: 81.7) spatial modes based on a singular value decomposition (SVD) of the outer-product of the leadfield matrix to retain > 92.8 % of the data variance. The projected data were further reduced to approximately 14-23 temporal modes (across-subjects mean: 19.7; with the maximum number of temporal modes set to 32) based on the SVD of the data covariance matrix.

Additionally, anatomical 'soft' priors (Litvak et al., 2011) were used that were defined by the AAL library (Tzourio-Mazoyer et al., 2002) using the MarsBaR toolbox (http://marsbar.sourceforge.net/) (Brett et al., 2002): Bilateral Heschl's gyri; Bilateral superior and inferior parietal cortex and bilateral superior and middle frontal gyri.

The inversion scheme calculated source time-courses at each vertex in the cortical mesh for each condition and subject. For statistical analysis, the average energy of the source time-course were computed over the entire 600 ms window for all frequencies between 2 and 30 Hz; the source energies were then interpolated into volumetric images in MNI space with 2 mm

voxels and spatially smoothed with a 12 mm FWHM isotropic Gaussian kernel. At the random effects or between-subject level, one source energy image per condition and subject was entered into paired t-tests. As in sensor space, we performed the following three statistical comparisons:

(1) The effect of *target detection* on processing of the first auditory tone: Hit1 vs. Miss1.

(2) The effect of *target detection* on processing the second auditory tone: Hit2 vs. Miss2.

(3) The interaction between *target detection* and *tone-position within a target pair*: Hit1-Miss1 vs. Hit2-Miss2.

To obtain coordinates for later source wave extraction, we additionally calculated the main effect of *target detection*: (Hit1 + Hit2) vs. (Miss1 + Miss2). The results were used for source wave extraction (see below).

2.7.4 Extracting source waveforms

Source waveforms were extracted as the first eigenvariate of all vertices being significantly different for hits relative to misses as determined by the main effect of target detection (± 2 mm) (**figure 3A**). As the polarity of the first eigenvariate is not uniquely defined for MEG data, we determined the polarity of the source time courses for each subject such that the consistency across participants was maximized (based on correlation and a 2^{nd} order singular value decomposition of the time courses across all participants). Please note that this procedure may potentially bias the extracted waveforms towards increased consistency. However, the ensuing source time-courses were used only for visualization (see figure 3), the Dynamic Causal Modelling analyses and the frequency analysis of the steady-state responses that do not depend on the polarity of the source waveforms. By contrast, the statistics was performed on data in sensor space or on the average energy of the source time-courses.

2.7.5 Steady-State responses (SSRs): Analysis and statistics

Steady-state responses were characterized only in source space, because the brief amplitude modulated tones elicited only weak steady-state activity that may evade sensor space analyses as a consequence of intersubject variability in neuroanatomy, positioning of the subject etc.. Furthermore, it is well established that auditory steady-state activity is generated in auditory cortices enabling us to impose priors on the source localization (Giani et al., 2012; Millman et al., 2010; Okamoto et al., 2010; Schoonhoven et al., 2003; Steinmann and Gutschalk, 2011).

To optimize model inversion selectively for localization of steady-state activity, we pre-processed the data differently for later characterization of steady-state responses in source space. In particular, the pre-processing was identical to the one described for event-related responses, except that the continuous data were low-pass filtered with a cut off at 100 Hz and epoched into segments from 0 ms to 300 ms (i.e., 300 ms length, including 12 cycles of the 40 Hz steady-state activity). Hence, the temporal window during source inversion was adjusted to a range from 0-300 ms. Moreover, since SSR have been shown to be localised within the auditory cortex, only bilateral Heschl's Gyri were used as a spatial prior. Otherwise the inversion was identical to the one described for event-related responses.

Before inversion, the data were projected to a subspace of 72-101 (across-subjects mean: 81.8) spatial modes based on a singular value decomposition (SVD) of the outer-product of the leadfield matrix to retain 95.2 % of the data variance. The projected data were further reduced to approximately 14-24 (across-subjects mean: 18.75) temporal modes (across subjects-mean) based on the SVD of the data variance matrix.

We extracted source-waveforms from bilateral auditory ROIs as described above. To estimate the amplitude at 40 Hz, we applied a fast Fourier Transform to the extracted time courses. First, we addressed the question if our stimuli evoked reliable SSRs. Therefore, we compared the amplitude at 40 Hz to the pooled amplitude of the adjacent sideband frequency. In paired t-tests we asked whether the amplitude was significantly greater at 40 Hz than at the sideband

frequencies. Next, we estimated the effect of target detection by subtracting the amplitude of misses from the amplitude of hits. In paired t-tests we evaluated whether the effect was significantly greater at 40 Hz than at sideband frequencies, separately for each target with the sequence.

2.7.6 Effective Connectivity Analysis: Dynamic Causal Modelling

Using dynamic causal modelling (DCM) we investigated how successful target detection is mediated via changes in effective connectivity between auditory and parietal areas. Since the awareness related negativity emerged only for the second tone within a target pair, DCM was only applied to the evoked responses of the second tone (i.e. Hit2 relative to Miss2) (David et al., 2005). As we had already localized the sources activated in our conditions via distributed source analysis, we performed DCM directly on the extracted sourcewaves after normalization using the LFP ('local field potential') option in SPM.

For each subject, 6 DCMs were constructed. Each DCM included the three regions where we obtained significant source activity across subjects: the left auditory cortex, the right auditory cortex and the right parietal cortex. The auditory inputs entered both auditory areas with a temporal onset of 100 ms (STD: 50) as soft prior.

The auditory cortices were bidirectionally connected via lateral connections. Furthermore, each auditory area was reciprocally connected with the parietal cortex via forward and backward connections. Holding the basic connectivity structure constant, we generated 2 x 3 = 6 DCMs by factorially manipulating which connections were modulated dynamically by auditory stream segregation and perceptual awareness. First, we manipulated whether the intrinsic connections within auditory cortices were modulated by 'successful target detection' (i.e. modulatory effect: present (i+) vs. absent (i-)). Second, we manipulated whether the effect of 'successful target detection' was associated with changes in the (1) forward (**Figure 4A, red arrows**), (2) backward (**Figure 4A, purple arrows**) or (3) forward and backward connections.

To assess the contributions of intrinsic, forward and backward connectivity to early and later awareness-related response differences, we estimated the 6 DCMs for source waves limited to early (0 – 300 ms) and early+late (from 0 – 550 ms) time windows (n.b. DCM can only be applied to evoked responses starting at stimulus onset; hence it is not possible to estimate a DCM that starts at 300ms)

To determine the most likely model given the observed data from all subjects, the 6 models were compared using Bayesian model selection separately for the early and early+late time windows. Bayesian model selection is based on the model evidence as approximated by the free energy that depends on both model fit and model complexity.

To avoid distortions by outlier subjects, Bayesian Model Selection was implemented in a random effects group analysis using a hierarchical Bayesian model that estimates the parameters of a Dirichlet distribution over the probabilities of all models considered. These probabilities define a multinomial distribution over model space enabling the computation of the posterior probability of each model given the data of all subjects and the models considered. To characterize our Bayesian Model Selection results at the random effects level, we report (i) the expectation of this posterior probability i.e. the expected likelihood of obtaining the k-th model for any randomly selected subject and (ii) the exceedance probability of one model being more likely than any other model tested (Stephan et al., 2009). The exceedance probability quantifies our belief about the posterior probability that is itself a random variable. Thus, in contrast to the expected posterior probability, the exceedance probability also depends on the confidence in the posterior probability.

Bayesian Model comparison enabled us to test the following two hypotheses. First, we expected that the earlier awareness related negativity (ARN) as an index of auditory stream segregation is caused by modulations of the intrinsic connections within auditory cortices (see Garrido et al., 2009; Kiebel et al., 2007). By contrast, we predicted that the later P300m awareness-related effects are mediated via recurrent processing between auditory and parietal cortices as

predicted by current theories of consciousness (Bekinschtein et al., 2009; Brancucci et al., 2011; Dehaene et al., 2006; Del Cul et al., 2007; Sergent et al., 2005; Shen and Alain, 2011) .

3. Results

3.1 Behaviour

Participants detected on average 39.67 ± 10% (mean ± standard deviation) and had 7.08 ± 4.73% false alarms, resulting in an across participants mean d' (i.e. d prime) of 1.32 ± 0.53 (across participants mean ± standard deviation). As previous studies suggested that the detection rate depends on target frequency (Bregman, 1990; Gutschalk et al., 2008), we also evaluated the detection rate separately across target frequencies. Indeed, a repeated measures ANOVA identified a significant effect of *target-frequency* on the detection rate ($F(1.93) = 15.77$, $p < 0.001$). Post hoc tests showed that the targets with the lowest target frequency (i.e.1222 Hz) were less often detected than all the other targets. Likewise, targets with a frequency of 1583 Hz were less often detected than targets at 2049 Hz (**table 1**).

Table 1 about here

3.2 Sensor space

In sensor space we investigated the effect of 'successful target detection' on auditory processing by comparing the activity for detected and undetected auditory targets in paired *t*-tests separately for target 1 and 2. Furthermore, we directly tested for the interaction between 'target detection' and *tone-position within a target pair*. Based on a priori hypotheses we evaluated the effect of 'target detection' in three time-windows of interest: (1) M_1: 40-90 ms, (2) ARN: 100-200ms, (3) M_3: 250-550 ms (**figure 2, table 2**) (Boutros and Belger, 1999; Gutschalk et al., 2008; Ishii et al., 2009; Konigs and Gutschalk, 2012; Wiegand and Gutschalk, 2012).

3.2.1 A neuronal processing cascade

Irrespective of participants' target detection all auditory targets evoked an early M_1 response with a bipolar topography over bilateral temporal sensors, i.e. suggesting that it forms the magnetic equivalent to the P50 (Boutros and Belger, 1999; Woldorff et al., 1993; Woldorff et al., 1998). However, for the first tone within a target pair only, we observed a significant difference in the M_1 component for hits and misses at ~60 ms (**figure 2A**). Further, we identified a significant *target detection* x *tone-position* interaction at 85 ms suggesting that the neural processes that enable target detection also depend on the target's sequential position.

In the mid-latency window, we observed a significant difference in neural activity only for the second target when comparing detected relative to undetected trials (**figure 2B**). When detected, the second target elicited an additional deflection with a dipolar topography opposite to the polarity of the M_1 component. In line with the previously reported so-called awareness related negativity (ARN) (Gutschalk et al., 2008; Konigs and Gutschalk, 2012; Wiegand and Gutschalk, 2012), the activity emerged at approximately 100 ms, peaked at ~ 150 ms and ended at ~200ms. Critically, as statistically confirmed by a significant *target detection* x *tone-position* interaction, this negativity was evident only for the second target, but not for the first target. This activity profile suggests that auditory stream segregation and perception of the sequential pair of tones induces this sustained negativity. By contrast, this negativity did not emerge for tone 1, even when participants detected the target pair and were aware of the presentation of tone 1. Thus, it is unlikely that this negativity is a necessary neural signature for detecting a single tone.

Detected relative to undetected targets evoked increased M_3 deflections (250 to 550 ms) (Bledowski et al., 2006; Ishii et al., 2009). This detection related effect was identified for both target 1 and target 2 (**figure 2 A & B**). More specifically, for target 1 activity differences for hit and misses were observed from 380 to 510 ms. For target 2, they peaked around 300 and 418

ms. Small differences in topography for hit1 and hit2 may perhaps indicate that the M3 response for tone 1 and the M3 response for tone 2 are generated by partly non-overlapping neural processes. However, the interaction between target detection and tone position was not statistically significant, so that we refrain from further discussions.

Figure 2 about here

3.2.2 Evaluation of the influence of target frequency on the sensor space results

At the behavioural level, the target detection rate significantly depended on carrier frequency of auditory target. This raises the question, whether the neural effects related to successful target detection may in fact reflect effects induced by different carrier frequencies. To evaluate this potential confound, we reanalyzed our data by sorting hits and misses into 5 new conditions according to the target's carrier frequency. At the random effects level, we then entered these data into a 2 (target detection: hits, misses) x 2 (target position: first, second) x 5 (carrier frequency: 1222, 1583, 2049, 2654 and 3437 Hz) repeated measure ANOVA. This ANOVA identified no significant effects of carrier frequency at $p < 0.05$ (corrected for multiple comparisons). Likewise, **figure 3C** shows that the effect of target detection was similar for all target frequencies.

Figure 3, Table 2 about here

3.3 Source space analysis

3.3.1 Event-related responses

To determine the neural generators related to successful target detection in sensor space, we performed source localization within a Bayesian framework (**table 3, figure 3**). The main effect of target detection revealed significant activity in Heschl's gyri bilaterally and in the right intraparietal cortices. These effects of target detection were also observed in right parietal cortex when testing for each tone of the target pair individually though at a more liberal threshold of 0.001 (uncorrected) (**figure 3A**).

Figure 3B shows the source-waves extracted from bilateral auditory and right parietal cortices for each condition. The sources of all three components of interest were localized at least in part within the auditory cortex. By contrast, late M3-like activity was predominantly generated by a source in the right parietal ROI. This is consistent with the well-established differentiation between the P3b and the P3a components of the P300. The parietal P3b component has been implicated in detection of rare events (e.g. in our case detection of a rare target pair in the context of multiple masking tones). By contrast, the frontal P3a component is mainly invoked for orienting responses such as shifting the attention to distractor items (Bledowski et al., 2004; Kok, 2001; Linden, 2005). As our paradigm places demands on target detection and does not require shifting of attention or complex contextual updates, the underlying neural sources may be predominantly located in parietal rather than frontal cortices.

Table 3 about here

3.3.2 Steady-State Responses

As expected, the 40 Hz amplitude modulated target tone evoked robust steady-state responses at the modulation frequency. The amplitude at 40 Hz was significantly increased relative to the sideband frequencies' amplitude.

However, our central question was whether this SSR differed depending on whether a target tone was successfully detected. Even though detected targets elicited stronger SSRs than misses, this increase in amplitude was not specific to 40 Hz frequency but also applied to sideband frequencies. The effect of target detection (Hit – Miss) was not significantly different for 40 Hz than for sideband frequencies (**table 4**). This null-result is consistent with the previous results by Gutschalk et al. (2008). However, null-results need to be interpreted with caution and cannot prove that the 40Hz SSR is not related to whether or not targets are segregated from the background mask. In fact, most recent studies (Ross et al., 2012) have suggested that the 40Hz SSR may consist of at least two oscillatory components with only one component being immune

to central masking and the other one being modulated by central masking. These results suggest that more finegrained analyses focusing on phase synchrony may potentially provide a different picture.

Table 4 about here

3.4 Effective connectivity: Dynamic Causal Modelling

Using DCM we investigated the contributions of intrinsic, forward and backward connections to modelling the difference in auditory and parietal source activity between Hits and Misses for early (<300ms) and early+late time windows.

Figure 4B shows the expected posterior probability of the 6 models in our 2 (modulatory effect on intrinsic connectivity: present vs. absent) x 3 (modulatory effect on: forward, backward or forward and backward extrinsic connectivity) factorial model space. For the early time window (<300 ms) that includes only the previously reported so-called 'awareness related negativity' but not the P300m component, the winning model included modulatory effects only for the forward connectivity and the intrinsic connectivity. Thus, modulation of the intrinsic connectivity is required to model the sustained negative ARN deflection suggesting that it may emerge via local dynamics within auditory association cortices.

By contrast, the feed-back model outperformed the feedforward model for the early+late time window that includes the P300m component as the most prominent sustained deflection. These results suggest that the later P300m may rely on top-down influences from parietal on auditory cortices. However, the difference in exceedance probability was less pronounced when comparing feedforward and feedback connections. To illustrate the influence of the modulations of the intrinsic and feedback connections to modelling the event-related responses, figure 4 shows the predictions of a model with (i) no modulations (i.e. predicted responses for detected and not detected tones are identical), (ii) only with modulations of intrinsic connections, (iii) with modulations of intrinsic and feedback connections together with the observed evoked source

responses.

These results largely converge with our hypotheses that the earlier so-called ARN relies on modulations of intrinsic connectivity within auditory cortices and the later P300m also on recurrent processing between parietal and auditory cortices. It is a little surprising that the winning model for the early+late tine window does not include modulations of both forward and backward connectivity. Potentially, this result may be explained by the fact that the forward-backward model incurs a higher model complexity penalty relative to the backward model.

Figure 4 about here

4. Discussion

The neural processes that are associated with auditory awareness have remained controversial. Using informational masking, recent MEG studies have linked perceptual awareness with a mid-latency component at about 150 ms that was coined the 'awareness related negativity' and localized in secondary auditory cortices (Gutschalk et al., 2008; Konigs and Gutschalk, 2012; Wiegand and Gutschalk, 2012). Yet, findings from mismatch negativity paradigms have associated auditory awareness with activations in a frontoparietal network (Bekinschtein et al., 2009; Diekhof et al., 2009). These discrepancies may be explained by the fact that previous analyses in informational masking focused predominantly on auditory cortices (Gutschalk et al., 2008) or pooled over multiple individual target tones. Yet, perceptual awareness in informational masking emerges in a complex neural processing cascade that builds up over multiple target tones.

Our results demonstrate that the neural responses for the first and second tones of the target are distinct. The first tone elicited a stronger early P50m and a slightly enhanced P300m response when participants successfully detected the target pair. By contrast, the second tone evoked the so-called awareness related negativity and a later strongly enhanced P300m

response (e.g.: Auksztulewicz et al., 2012; Bekinschtein et al., 2009; Del Cul et al., 2007; Hillyard et al., 1971; Ishii et al., 2009; Kok, 2001; Sergent et al., 2005; Shen and Alain, 2011; van Aalderen-Smeets et al., 2006). This neural response pattern challenges conclusions drawn from previous studies in several aspects (Konigs and Gutschalk, 2012).

First, our results demonstrate that differences for detected and undetected targets emerge already at 50 ms poststimulus for the first tone within a target pair (i.e. the P50m component). Potentially, the 2nd tone also elicits an enhanced P50m for detected tones, but this small effect might be swamped by the larger effect of the later sustained negativity (see figure 2). This very early P50m enhancement may in part result from participants spontaneously attending to the relevant target frequency band (see Guterman et al., 1992; Karns and Knight, 2009; for attentional modulation of the P50 component). The enhanced P50m may reflect a stronger auditory representation of the first tone that in turn facilitates processing of the second tone.

Surprisingly, the previously described awareness related negativity (ARN) is elicited only for the 2nd but not the 1st tone. As participants must have been aware of both tones for successful target pair detection, these findings question the previously proposed critical role of the ARN in detection and awareness of single tones (Gutschalk et al., 2008). Instead, they suggest that this negativity may be a neural signature of segregating the target pair from the auditory background mask, which can only emerge after the presentation of the 2nd tone. Since the target tones were separated from the masking tones via a frequency gap in the current paradigm, we expect that they interact predominantly at later processing stages in belt or parabelt areas rather than already at the basilar membrane (for review and further discussion see (Gutschalk and Dykstra, 2014)).

While the enhanced M50 was observed only for the 1st tone, an increased P300m component generated predominantly by parietal cortices was observed for both tones, when they were detected. As the P300m amplification is the only neural feature that was present for both target tones when they were detected, later processing in parietal cortices may play a critical role in

successful detection performance and thus auditory awareness (e.g.: Auksztulewicz et al., 2012; Bekinschtein et al., 2009; Del Cul et al., 2007; Hillyard et al., 1971; Ishii et al., 2009; Kok, 2001; Sergent et al., 2005; Shen and Alain, 2011; van Aalderen-Smeets et al., 2006). This conjecture converges with recent EEG studies showing an enhanced P300 component for perceptual awareness of global auditory regularities in mismatch negativity paradigms (Bekinschtein et al., 2009). Furthermore, it dovetails nicely with studies in the visual domain showing an enhanced P300 for conscious access in attentional blink paradigms (Sergent et al., 2005).

Collectively, our results suggest that the masker tones interact with target processing at multiple processing stages in primary/higher order auditory and parietal cortices as indicated by differences in the P1m, ARN and P300m for detected vs. undetected targets. In primary auditory cortices informational masking may alter the sensory representations of individual target tones potentially via saliency or attentional mechanisms (see Elhilali et al., 2009). Next, in higher order auditory cortices (e.g. belt and parabelt areas) auditory target object representations may become segregated from complex background masks and potentially further stabilized via top-down influences from parietal cortices.

Using Dynamic Causal Modelling we then investigated the effective connectivity changes within and between auditory and parietal cortices that are associated with auditory stream segregation and awareness. In particular, we asked whether perceptual awareness relies on recurrent processing between auditory and parietal areas (Auksztulewicz et al., 2012; Dehaene and Changeux, 2011; Dehaene et al., 2006; Lamme, 2006; Lamme and Roelfsema, 2000; Ro et al., 2003). Bayesian model comparison demonstrated that the ARN relies critically on local connectivity changes within auditory cortices. This converges with the notion that intrinsic connections play a critical role in perceptual grouping (Roelfsema, 2006) and generation of auditory mismatch responses (Garrido et al., 2008; Garrido et al., 2009; Kiebel et al., 2007).

By contrast, neural activity occurring later up to 600 ms poststimulus was better modelled by a

DCM that accommodated modulations of intrinsic and backward connections. Comparison of the predicted and observed source timecourses revealed that the backwards connections were important for modelling the later P300m component. Collectively, these results suggest that auditory awareness in informational masking is associated with recurrent loops and top-down influences from parietal to auditory cortices (Garrido et al., 2007). This is in line with a recent DCM study showing a disruption of top-down processes in patients in vegetative state relative to normal controls in a recent mismatch negativity paradigm (Boly et al., 2011).

In conclusion, our findings reveal that auditory stream segregation and awareness in informational masking relies on a complex processing cascade. The previously reported ARN was observed only for the second tone when the target pair could be segmented from the background mask pointing towards a role in auditory figure ground segregation during informational masking. Dynamic causal modelling suggested that these processes rely on dynamic changes in local connectivity within auditory cortices. By contrast, a later sustained P300m response was observed jointly for both tones when they were consciously perceived and successfully detected. These later awareness-related P300m responses were mediated via recurrent connectivity between auditory and higher order association cortices pointing towards a critical role in auditory awareness.

5. References

Allen, J., Kraus, N., Bradlow, A., 2000. Neural representation of consciously imperceptible speech sound differences. Percept Psychophys 62, 1383-1393.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839-851.

Auksztulewicz, R., Spitzer, B., Blankenburg, F., 2012. Recurrent neural processing and somatosensory awareness. J Neurosci 32, 799-805.

Baars, B.J., 1993. A cognitive theory of consciousness. Cambridge University Press.

Backer, K.C., Alain, C., 2012. Orienting attention to sound object representations attenuates change deafness. Journal of experimental psychology: human perception and performance 38, 1554-1566.

Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., Naccache, L., 2009. Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci U S A 106, 1672-1677.

Bledowski, C., Cohen Kadosh, K., Wibral, M., Rahm, B., Bittner, R.A., Hoechstetter, K., Scherg, M., Maurer, K., Goebel, R., Linden, D.E., 2006. Mental chronometry of working memory retrieval: a combined functional magnetic resonance imaging and event-related potentials approach. J Neurosci 26, 821-829.

Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., Linden, D.E., 2004. Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study. J Neurosci 24, 9353-9360.

Boly, M., Garrido, M.I., Gosseries, O., Bruno, M.A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S., Friston, K., 2011. Preserved feedforward but impaired top-down processes in the vegetative state. Science 332, 858-862.

Boutros, N.N., Belger, A., 1999. Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. Biol Psychiatry 45, 917-922.

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat Vis 10, 433-436.

Brancucci, A., Franciotti, R., D'Anselmo, A., Della Penna, S., Tommasi, L., 2011. The sound of consciousness: neural underpinnings of auditory perception. J Neurosci 31, 16611-16618.

Bregman, A.S., 1990. Auditory scene analysis: The perceptual organization of sound. MIT press, Cambridge, MA.

Brett, M., Anton, J., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox. Neuroimage 16, Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan.

Cusack, R., 2005. The intraparietal sulcus and perceptual organization. J Cogn Neurosci 17, 641-651.

David, O., Harrison, L., Friston, K.J., 2005. Modelling event-related responses in the brain. Neuroimage 25, 756-770.

Dehaene, S., Changeux, J.P., 2011. Experimental and theoretical approaches to conscious processing. Neuron 70, 200-227.

Dehaene, S., Changeux, J.P., Naccache, L., Sackur, J., Sergent, C., 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends Cogn Sci 10, 204-211.

Dehaene, S., Charles, L., King, J.-R., Marti, S., 2014. Toward a computational theory of conscious processing. Current opinion in neurobiology 25, 76-84.

Dehaene, S., Sergent, C., Changeux, J.-P., 2003. A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proceedings of the National Academy of Sciences 100, 8520-8525.

Del Cul, A., Baillet, S., Dehaene, S., 2007. Brain dynamics underlying the nonlinear threshold for access to consciousness. PLoS Biol 5, e260.

Diekhof, E.K., Biedermann, F., Ruebsamen, R., Gruber, O., 2009. Top-down and bottom-up modulation of brain structures involved in auditory discrimination. Brain Res 1297, 118-123.

Elhilali, M., Xiang, J., Shamma, S.A., Simon, J.Z., 2009. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. PLoS biology 7, e1000129.

Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Kilner, J.M., 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42, 936-944.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Friston, K.J., 2007. Evoked brain responses are generated by feedback loops. Proc Natl Acad Sci U S A 104, 20961-20966.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Friston, K.J., 2009. Dynamic causal modeling of the response to frequency deviants. J Neurophysiol 101, 2620-2631.

Giani, A.S., Ortiz, E., Belardinelli, P., Kleiner, M., Preissl, H., Noppeney, U., 2012. Steady-state responses in MEG demonstrate information integration within but not across the auditory and visual senses. Neuroimage 60, 1478-1489.

Guterman, Y., Josiassen, R.C., Bashore, T.R., Jr., 1992. Attentional influence on the P50 component of the auditory event-related brain potential. Int J Psychophysiol 12, 197-209.

Gutschalk, A., Dykstra, A.R., 2014. Functional imaging of auditory scene analysis. Hearing Research 307, 98-110.

Gutschalk, A., Micheyl, C., Oxenham, A.J., 2008. Neural correlates of auditory perceptual awareness under informational masking. PLoS Biol 6, e138.

Haynes, J.D., Driver, J., Rees, G., 2005. Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. Neuron 46, 811-821.

Hillyard, S.A., Squires, K.C., Bauer, J.W., Lindsay, P.H., 1971. Evoked potential correlates of

auditory signal detection. Science 172, 1357-1360.

Ishii, R., Canuet, L., Herdman, A., Gunji, A., Iwase, M., Takahashi, H., Nakahachi, T., Hirata, M., Robinson, S.E., Pantev, C., Takeda, M., 2009. Cortical oscillatory power changes during auditory oddball task revealed by spatially filtered magnetoencephalography. Clin Neurophysiol 120, 497-504.

Karns, C.M., Knight, R.T., 2009. Intermodal auditory, visual, and tactile attention modulates early stages of neural processing. J Cogn Neurosci 21, 669-683.

Kidd, G., Jr., Mason, C.R., Deliwala, P.S., Woods, W.S., Colburn, H.S., 1994. Reducing informational masking by sound segregation. J Acoust Soc Am 95, 3475-3480.

Kidd, G., Jr., Mason, C.R., Richards, V.M., 2003. Multiple bursts, multiple looks, and stream coherence in the release from informational masking. J Acoust Soc Am 114, 2835-2845.

Kiebel, S.J., Garrido, M.I., Friston, K.J., 2007. Dynamic causal modelling of evoked responses: the role of intrinsic connections. Neuroimage 36, 332-345.

Kiebel, S.J., Garrido, M.I., Moran, R., Chen, C.C., Friston, K.J., 2009. Dynamic causal modeling for EEG and MEG. Hum Brain Mapp 30, 1866-1876.

Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3? Perception 36.

Koch, C., Tsuchiya, N., 2007. Attention and consciousness: two distinct brain processes. Trends in cognitive sciences 11, 16-22.

Kok, A., 2001. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38, 557-577.

Konigs, L., Gutschalk, A., 2012. Functional lateralization in auditory cortex under informational masking and in silence. Eur J Neurosci.

Lamme, V.A., 2006. Towards a true neural stance on consciousness. Trends Cogn Sci 10, 494-501.

Lamme, V.A., Roelfsema, P.R., 2000. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 23, 571-579.

Linden, D.E., 2005. The p300: where in the brain is it produced and what does it tell us? Neuroscientist 11, 563-576.

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., Friston, K., 2011. EEG and MEG data analysis in SPM8. Comput Intell Neurosci 2011, 852961.

Micheyl, C., Shamma, S.A., Oxenham, A.J., 2007. Hearing out repeating elements in randomly varying multitone sequences: a case of streaming? Hearing–From Sensory Processing to Perception, 267-274.

Millman, R.E., Prendergast, G., Kitterick, P.T., Woods, W.P., Green, G.G., 2010. Spatiotemporal reconstruction of the auditory steady-state response to frequency modulation using magnetoencephalography. Neuroimage 49, 745-758.

Moore, C.J.M., 1995. Frequency analysis and masking. In: Moore, C.J.M. (Ed.), Hearing. Academic Press, Inc., USA, pp. 161-205.

Neff, D.L., Green, D.M., 1987. Masking produced by spectral uncertainty with multicomponent maskers. Percept Psychophys 41, 409-415.

Okamoto, H., Stracke, H., Bermudez, P., Pantev, C., 2010. Sound Processing Hierarchy within Human Auditory Cortex. J Cogn Neurosci.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011, 156869.

Picton, T.W., John, M.S., Dimitrijevic, A., Purcell, D., 2003. Human auditory steady-state responses. Int J Audiol 42, 177-219.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N.S., Lane, D., 2003. Feedback contributions to visual awareness in human occipital cortex. Curr Biol 13, 1038-1041.

Roelfsema, P.R., 2006. Cortical algorithms for perceptual grouping. Annu Rev Neurosci 29, 203-227.

Ross, B., Miyazaki, T., Fujioka, T., 2012. Interference in dichotic listening: the effect of contralateral noise on oscillatory brain networks. European Journal of Neuroscience 35, 106-118.

Sadaghiani, S., Hesselmann, G., Kleinschmidt, A., 2009. Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. J Neurosci 29, 13410-13417.

Schoonhoven, R., Boden, C.J., Verbunt, J.P., de Munck, J.C., 2003. A whole head MEG study of the amplitude-modulation-following response: phase coherence, group delay and dipole source analysis. Clin Neurophysiol 114, 2096-2106.

Sergent, C., Baillet, S., Dehaene, S., 2005. Timing of the brain events underlying access to consciousness during the attentional blink. Nat Neurosci 8, 1391-1400.

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., Tallon-Baudry, C., 2013. Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. Current Biology 23, 150-155.

Shen, D., Alain, C., 2011. Temporal attention facilitates short-term consolidation during a rapid serial auditory presentation task. Exp Brain Res 215, 285-292.

Steinmann, I., Gutschalk, A., 2011. Potential fMRI correlates of 40-Hz phase locking in primary auditory cortex, thalamus and midbrain. Neuroimage 54, 495-504.

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., Griffiths, T.D., 2011. Brain bases for auditory stimulus-driven figure-ground segregation. J Neurosci 31, 164-171.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a

macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273-289.

van Aalderen-Smeets, S.I., Oostenveld, R., Schwarzbach, J., 2006. Investigating neurophysiological correlates of metacontrast masking with magnetoencephalography. Advances in Cognitive Psychology 2, 21-35.

Wiegand, K., Gutschalk, A., 2012. Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. Neuroimage 61, 62-69.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., Bloom, F.E., 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc Natl Acad Sci U S A 90, 8722-8726.

Woldorff, M.G., Hillyard, S.A., Gallen, C.C., Hampson, S.R., Bloom, F.E., 1998. Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. Psychophysiology 35, 283-292.

6. Table and figure legends

Table 1. Behavioural results: Accuracy is shown separately for target present trials (i.e. % hits) and target absent trials (i.e. % correct rejections) X the 5 different target frequencies.

Table 2. Sensor space analysis - Statistical results. Image coordinates (mm*mm*ms) are given

for the peak voxel. Sensor locations correspond to CTF conventions (first letter: R = right and L = left; second letter: T = temporal, O = occipital, P = parietal, F = frontal; numbering: 1st digit = row and 2nd digit = column). P-values are corrected for multiple comparisons for all voxels within the time window of interest.

Table 3. Sensor space analysis - Statistical results: MNI co-ordinates are provided for the peak p-value corrected for multiple comparisons within the specific search volume of interest encompassing left and right Heschl's Gyrus, left and right superior and inferior parietal cortex, superior and middle frontal gyri.

Table 4. Steady-state responses - Statistical results. HG = Heschl's Gyrus

Figure 1. Informational masking paradigm. (A) Schematic representation of an example trial. Participants detected a pair of two sequentially presented tones, i.e. the target (black) embedded within a multi-tone mask (grey). The two tones were presented with a fixed interstimulus interval in a protected region in frequency space (grey shading). After each trial participants indicated whether they detected a target pair. (B) Experimental conditions of interest. The neural responses were analysed separately for target 1 and 2 within each target pair. Hence, our 2 x 2 experimental design factorially manipulated: (1) awareness (hits, misses) and (2) target position (first, second).

Figure 2. Sensor space results. (***A***) Top/Bottom: Statistical F-maps, topography of amplitude of

evoked responses for hits and misses are shown separately for the first tone within the target pair at latencies where the difference in response amplitude between hits and miss were significant. Middle: Butterfly plot of the activity evoked by the first tone separately for hits (dark blue) and misses (red). (***B***) Top/Bottom: Statistical F-maps, topography of amplitude of evoked responses for hits and misses are shown separately for the second tone within the target pair at latencies where the difference in response amplitude between hits and miss were significant. Middle: Butterfly plot of the activity evoked by the second tone separately for hits (light blue) and misses (orange). (***C***) Butterfly plots of the evoked activities for the first (left column) and second (right column) target tones, are shown separately for each target frequency.

Figure 3. Source space results. (***A***) Main effect of awareness. Statistical results are displayed on a canonical, inflated MNI brain. (***B***) Extracted source waves. Source waves of hit1 (dark blue), hit2 (light blue), Miss1 (red) and Miss2 (orange) are displayed together with shaded error bars for the left auditory (left column) right auditory (middle column) and parietal cortical region of interest (right column). Vertices that are selected for source wave extraction are represented as red dots on a cortical mesh of a representative participant.

Figure 4. Dynamic causal modelling (DCM). (***A***) Six candidate DCMs were generated by factorially manipulating whether awareness modulates (1) extrinsic connections (i.e. feedforward, feedback or both) and (2) intrinsic connections within the auditory cortex (i.e. present vs. absent). (***B, C***) Bayesian model comparison (random effects analysis) for early (< 300 ms, upper row) and early+late neural dynamics (0 – 550 ms, lower row). The matrix shows the expected posterior probability (B) and exceedance probability of the six DCMs in a factorial fashion. (***D***) The observed (dashed line) source activity is shown together with source activity

predicted by the early+late model (i) with no modulatory effects (grey solid), (ii) with modulatory effects on intrinsic connections alone (dotted orange or blue) and (iii) with modulatory effects on intrinsic and feedback connections (i.e. the winning model, solid) for hit2 (light blue solid) and miss2 (orange solid) conditions.

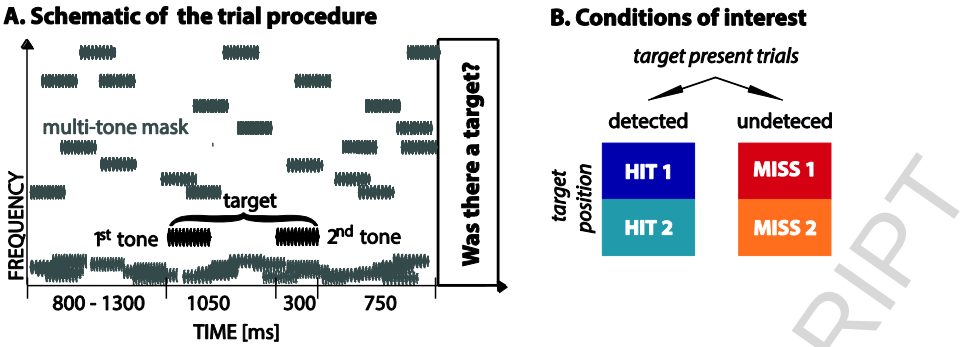

Figure 1

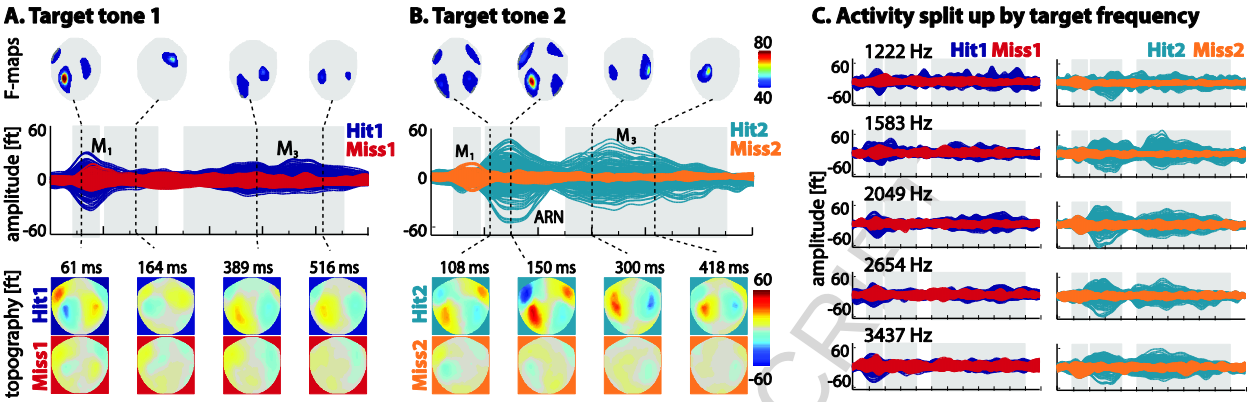

Figure 2

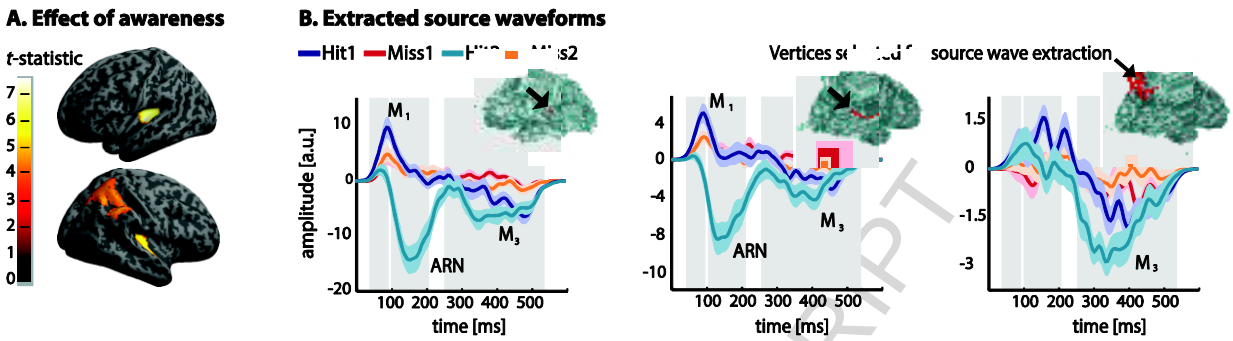

Figure 3

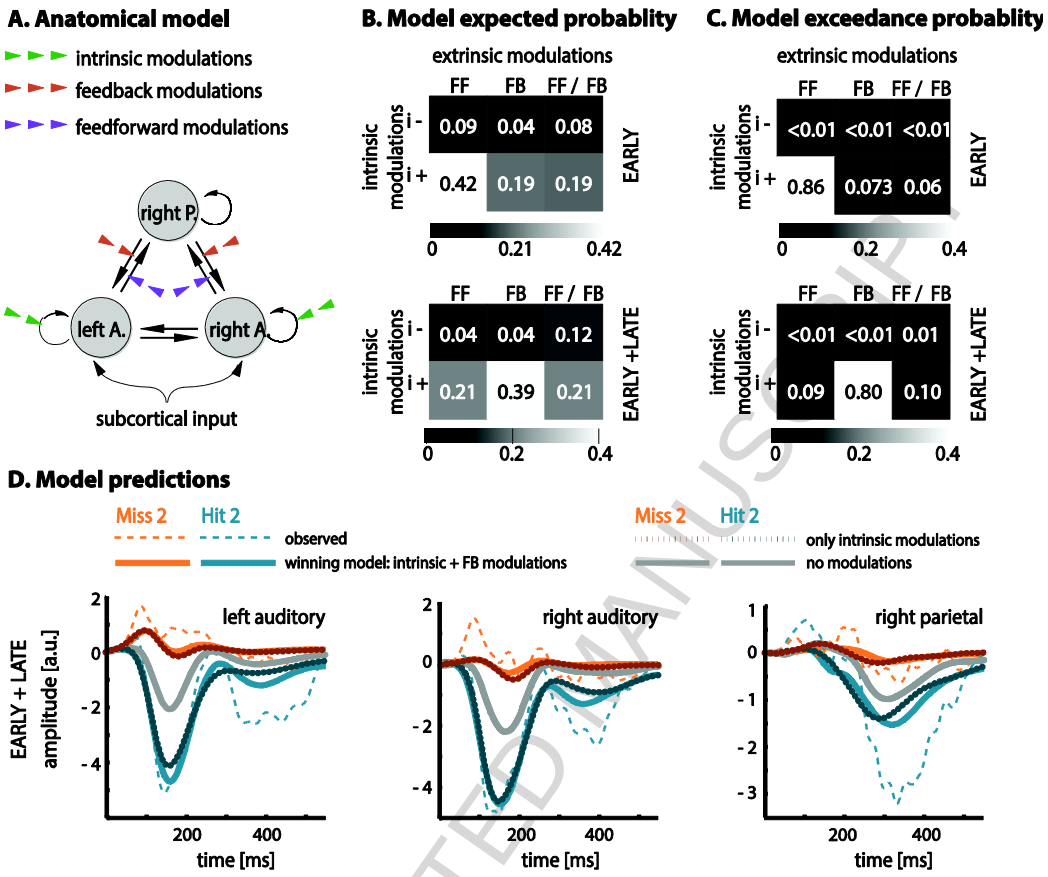

Figure 4

Table 1. Behavioural results: Accuracy for different target frequencies

	Target frequency				
	1222 Hz	**1583 Hz**	**2049 Hz**	**2654 Hz**	**3437 Hz**
	15.35 ±	35.07±	51.98±	48.76±	45.92 ±
% Hit	7.50%	17.25%	23.18%	19.21%	8.88%
% Corr.	90.55 ±	90.79 ±	91.53 ±	92.79 ±	93.44 ±
rejection	6.65%	7.31%	6.05%	6.70%	5.42%

Table 2. Sensor space analysis: Statistical results

Coordinates			Nearest	Z-score	p-value
mm	mm	ms	**sensor**	**(peak)**	**(corrected)**
Simple main effects of awareness					
Target 1: Hit1 ≠ Miss 1					
-32	-38	61	LT16	5.65	< 0.001
30	16	164	RF55	4.98	0.007
28	-17	389	RC25	4.74	0.018
-30	-33	516	LP55	4.54	0.039
Target 2: Hit2 ≠ Miss 2					
-30	-46	150	LT27	5.45	0.001
-17	-17	418	LP23	5.19	0.002
55	-17	300	RT34	5.06	0.004
-53	21	160	LT33	4.75	0.014
36	-38	108	RT16	4.52	0.035
Interaction: Awareness x target-position					
(Hit1 – Miss1) ≠ (Hit2 – Miss2)					
-34	-38	155	MLT16	5.36	0.001
-38	-46	85	MLT37	4.99	0.006

Table 3. Source space analysis: Statistical results

Brain Region	MNI Coordinates			Cluster size	Z-score (peak)	p-Value (corrected)
	x	y	z			
Simple main effect of awareness						
Target 1: Hit1 > Miss1						
R. transverse temporal gyrus (HG)	50	-12	4	155	4.54	0.016
L. transverse temporal gyrus (HG)	-46	-18	6	124	4.53	0.017
R. inferior parietal cortex	48	-46	50	982	3.69	0.264
Target 2: Hit2 > Miss 2						
L. transverse temporal gyrus (HG)	-42	-20	6	123	4.77	0.004
R. transverse temporal gyrus (HG)	48	-16	6	138	3.98	0.079
R. intraparietal sulcus	46	-46	50	91	3.30	0.479
R. transverse parietal sulcus	30	-56	50	37	3.26	0.512
R. superior parietal gyrus	40	-46	58	15	3.18	0.593
R. intraparietal sulcus	34	-40	52	1	3.10	0.985
Interaction: Awareness x target-position						
(Hit1 – Miss1) ≠ (Hit2 – Miss2)						
L. transverse temporal gyrus (HG)	-44	-20	6	93	3.40	0.109
Main effect of awareness						
(Hit1 + Hit2) ≠ (Miss1 + Miss2						

L. transverse temporal gyrus (HG)	-44	-20	6	127	5.07	< 0.001
R. transverse temporal gyrus (HG)	46	-16	6	151	4.45	0.002
R. inferior parietal cortex	48	-46	50	1105	3.85	0.022

Table 4. Steady state responses: statistical results.

	Target 1		Target 2	
	Hit 1	***Miss 2***	***Hit2***	***Miss2***
Amplitude at 40 Hz ≠ sideband frequency				
Left HG	$t(19) = 2.70$, $p = 0.014$	$t(19) = 4.80$, $p < 0.001$	$t(19) = 2.82$, $p < 0.011$	$t(19) = 4.30$, $p < 0.001$
Right HG	$t(19) = 4.15$, $p < 0.001$	$t(19) = 5.35$, $p < 0.001$	$t(19) = 4.47$, $p < 0.001$	$t(19) = 4.72$, $p < 0.001$
Hit – Miss for Amplitude at 40 Hz ≠ sideband frequency				
Left HG	$t(19) = -0.19$, $p = 0.852$		$t(19) = 0.65$, $p = 0.525$	
Right HG	$t(19) = 0.20$, $p = 0.847$		$t(19) = 1.16$, $p = 0.262$	

Highlights

- Informational masking enables modulation of auditory awareness.
- Auditory awareness evolves in a neural processing cascade of M50, ARN and M300.
- ARN mediates stream segregation via intrinsic connectivity within auditory cortices.
- M300 as a signature of auditory awareness relies on parieto-temporal interactions.
- Awareness emerges via recurrent processing between parietal and auditory cortices.